SOME NOTES ON THE EVOLUTION OF THE GENERAL LINEAR MODEL

by

R. P. Kershner
Arnar-Stone Laboratories
McGaw Park
Waukegan, Illinois, 60085

BU-678-M June 1979

Abstract

Following the development of the theory of least squares, three key steps subsequently led to the development of the general linear model in its present form. The three steps are: 1) work in conceptualizing and refining the ideas of correlation which took place in the latter part of the ninteenth century, 2) development of the analysis of variance by Fisher, and 3) work in the 1930's and 1940's by researchers attempting to mathematize and generalize Fisher's thoughts on the partitioning of sources of variation. This paper discusses these three steps and provides a broad view of the historical development of the general linear model from Galton's ideas on correlation to the present day.

Paper No. BU-678-M in the Biometrics Unit Mimeo Series, Cornell University, Ithaca, New York.

1. Introduction

Of all the tools in the statistician's toolbox, none is more widely used than the theory and application of the general linear model. The range of application of this theory is extremely broad and includes, for example, the analysis of designed experiments, the analysis of covariance, curve fitting and prediction, analysis of survey data, and variance components estimation.

The present form of the general linear model theory has been abstracted, refined, and generalized considerably from the early tentative use first visualized by Fisher (e.g., Fisher, 1915 and 1921). Recently, authors such as Hocking and Speed (1975), Speed and Hocking (1976), Speed, Hocking and Hackney (1978), and Urquhart and Weeks (1979) are advocating a much simplier formulation of the general linear model by expressing models in terms of the cell means. In this way estimation of effects and tests of hypotheses may be formulated from an empirical rather than from a strictly theoretical point of view. In this respect this is a return to the form used 40 or more years ago. In returning to the earlier and perhaps more heuristic formulation, the methods have been richly embellished by the general theoretical concepts which have developed since the early papers of Fisher.

The purpose of this paper is to trace the development of the general linear model. This is obviously an extremely broad and many-faceted topic. Any attempt to give a detailed account of every aspect of the historical development of the general linear model within a paper of this size is bound to meet with failure — a lesson only learned from hindsight. Various chapters of this story have been given elsewhere, and this is yet another attempt to interrelate some of the developments that have taken place in the past. This paper will focus on the following: 1) the development of correlation beginning in the mid-1800's to its refinement as an intraclass correlation model, 2) trace Fisher's contribution through his development of the analysis of variance, and 3) discuss some of the work in the 1930's and 1940's which resulted in establishment of the general linear model along with its geometrical abstraction.

The spirit of this paper is such that we will occasionally take time to pause and reflect on some aspects of histrocial interest to present-day statistics which serve to highlight the remarkable tenacity of some of the early workers in statistics. In particular, one never ceases to marvel at the remarkable ingenuity of Galton, who, in 1885, deduced the form of the bivariate normal surface from some simple data smoothing techniques.

We are not covering the discovery and development of the theory of least squares during the ninteenth century, as this topic has been extensively covered in the past. The interested reader is referred to the papers of Plackett (1972), Eisenhart (1964), Seal (1967), and Anderson (1978).

2. The Development of Correlation

2.1. Introduction

The concept of correlation quite naturally arises when considering the joint behavior of two or more random variables. Thus some of the earliest work in correlation was linked to the investigations of the behavior of multivariate random variables. Galton stands apart from many of the other workers, such as Bravis, in light of his particularly pragmatic approach to the investigation of correlation.

2.2. The multivariate normal distribution

Bravis (1846) presented the first formulation for the multivariate normal surface for 2 and 3 variables. His formulation was in terms of random variables X_i with errors defined by

$$\epsilon_i = X_i - E(X_i) = \Sigma\, \alpha_i \epsilon_i \tag{1}$$

where ϵ_i are iid. $N(0,\sigma_i^2)$ and the α_i's and σ_i^2's are assumed to be known. Seal (1967) noted that these assumptions, together with the absence of any practical applications lead Pearson (1920) to conclude that Galton was the first to consider practical applications of the correlation between two directly observable variables. Specifically, Pearson's (1920) objection to the early workers such as Bravis was that he thought that they were dealing with indirectly observed variables which were linear functions of the directly observed but independent variables on the right-hand side of (1). He notes (1920) ". . . that the directly measured quantities might themselves be correlated does not seem to have occurred to the many writers on the theory of errors.". Seal (1967) noted that in this respect Pearson was being less than just to Schols (1875) who discussed the application of the bivariate normal to artillery fire at a target and critized those who assumed that errors in the horizontal and vertical direction were independent. It should be noted that earlier Pearson (1895) seemed more willing to acknowledge Bravis's contribution.

Some generalizations of the multivariate normal distribution were given by Edgeworth (1892), who extended the multivariate distribution to four variables, and Pearson (1896), who provided the complete generalization to p variables.

In contrast to these theoretical workers, Galton was cast in more of a practical mold. His contribution was in providing the basic concepts of correlation and regression and his ingenious use of simple data-analytic tools in arriving at a means of estimating the correlation coefficient from small samples. More importantly, his work provided the inspiration for other workers, most notably Weldon, Edgeworth, and Karl Pearson and his co-workers.

2.2 Francis Galton

Galton's work on practical applications of the correlation coefficient and his introduction of the concept of regression had a significant impact on statistics. The appearance of Natural Inheritance in 1889 inspired the likes of Pearson, Weldon and Edgeworth who, along with Pearson's co-workers such as Yule, extended Galton's ideas on correlation. Further generalizations were given by Fisher (1915, 1918) who considered the intraclass correlation model. Out of this preliminary work grew the analysis of variance.

At an early stage in his research, Galton noted

> "the curious regularity commonly observed in the statistical peculiarities of great populations during a long series of generations. The large do not always beget the large, nor the small the small, and yet the observed proportions between the large and the small in each degree of size and in every quality, hardly varies from one generation to another." (Galton (1889))

He notices these trends earlier when investigating the sizes of several generations of sweet pea seeds (Galton (1877)). This tendency was termed reversion (now known as regression) and he (Galton (1877)) defined it as

> ". . . the tendency of the ideal mean filial type to depart from the parental type, reverting to what may be roughly and perhaps fairly described as the average ancestrial type. If family variability had been the only process in simple descent that affected the characteristics of a sample the dispersion of

> the race from its mean ideal type would indefinitely increase with the number of generations, but reversion checks this increase, and brings it to a standstill."

In 1877 Galton plotted the average diameters of sweet pea seeds from offspring versus those of their parents. A straight line was then drawn through the data by eye and the slope was calculated as approximately one-third. This, as noted by Pearson (1930, Chapter 14), was the first regression line.

In 1885 Galton investigated the inheritance of stature of adult parents and their offspring. His data consisted of a cross-tabulation of parential heights with children heights. The original observations were reduced to deviations from the medians and were then expressed as unit deviations in terms of their respective quartiles. The resulting tabulation is reproduced in Pearson (1920). Galton further smoothed the data by calculating medians of four adjacent cells and connected like observations by drawing the elipsoidal curves of "countours of equal frequency". Galton (1885) noted that " . . . the original data ran somewhat roughly and I had to smooth them with tender caution.". Beginning with this simple cross-tabulation, Galton, through a leap of insight, then observed that the underlying distribution was bivariate normal and that the marginal distributions were normal with equal variance reduced by a constant factor which was proportional to the ratios of the variances in the parent population. He then deduced the positions of the regression lines which corresponded to "conjugate diameters of the variate axis". In order to arrive at the precise mathematical representation, Galton consulted J. D. H. Dixson, a mathematician at Cambridge, who easily derived the correct analytic formulation of a bivariate normal density.

Galton (1888) introduced the term correlation for the first time: ". . . two variable organis are . . . correlated when the variation of one is accompanied on the average by more or less variation of the other, and in the same direction.". The obvious generalization of this definition to include negative correlation was provided by Weldon (1892).

The early correlation coefficients were called "indices of correlation" by Galton and were calculated from a data analytic approach using medians and quartile deviations. A straight line was drawn through the smoothed arrays and its slope was read off as the correlation coefficient. In using this technique Galton was operating under assumptions of normality and homoschesticity. He was, in fact, considering models of the form $\underset{\sim}{y} = b_1 \underset{\sim}{x}$ where $\underset{\sim}{y}$ would be a vector of observations for parents and $\underset{\sim}{x}$ the corresponding observations for the offspring. Under such a model it follows that $r = b_1$, i.e., the correlation coefficient and the slope of the regression equation are the same.

Weldon (1889, 1890) used Galton's methods to calculate correlation coefficients for several races of the same species of shrimp and observed that the values tended to cluster about a common value. Correlation coefficients were compared on a strictly subjective basis since the distributional properties were still unknown. Weldon (1889) called the correlation coefficients "Galton Functions" in honor of their founder. It was Edgeworth (1892) who was responsible for coining the term "coefficient of correlation".

A very significant step in the derivation of the correlation coefficient was taken by Yule (1897) who explicitly showed the relationship between least squares regression and correlation. In this same paper he introduced the partial correlation coefficient, which has previously been referred to as the "net" correlation.

By modern standards, correlation has taken a back-seat to the much more powerful methods embodied within the framework of the general linear model.

However, it did provide the necessary framework for Fisher. Galton's work on correlation must be viewed in the spirit of the times. Correlation was the first quantification of biological data, and one can readily grasp the excitement generated by this new concept as expressed by Galton in his introduction to Natural Inheritance:

> ". . . those who care to brace themselves to a sustained effort, need not feel much regret that the road to be travelled over is indirect, and does not admit of being mapped beforehand in a way they can clearly understand. It is full of interest of its own. It familiarizes us with the measurement of variability, and with curious laws of chance that apply to a vast diversity of social subjects. This part of the inquiry may be said to run along a road on a high level, that affords wide views in unexpected directions, and from which easy descents may be made to totally different goals to those we have now to reach. I have a great subject to write upon . . ."

3. R. A. Fisher and the Analysis of Variance

3.1. Introduction

Certainly without question R. A. Fisher is the father of modern statistics. His contribution was marvelously rich and many of the ideas which he originated read like a syllabus of a first year's course in statistics; e.g., the likelihood function and maximum likelihood estimation; the concepts of sufficiency, efficiency, completeness; the analysis of variance; many small sample testing procedures such as testing the significance of regression coefficients; the Z distribution (later called F in honor of its founder by Snedecor); and an introduction to basic experimental design including ideas on factorial experiments, the use of the Latin square, confounding, and concepts of randomization.

In sketching Fisher's development of the analysis of variance it is tempting to take a look at the origin of the basic principles of experimental design as originally put forth by Fisher. This will be avoided for the most part, but some references to early design work will be noted where they directly impact on Fisher's maturing view of the analysis of variance.

Fisher made significant contributions to both statistics and genetics. Statisticians are often surprised to learn that he contributed so much to genetics while geneticists are equally surprised to learn about his massive contributions to statistics. He developed the ideas of blood grouping during World War II and unraveled the complexities of the Resus system, which is responsible for erythroblastosis foetalis.

In 1919 Fisher accepted a post as statistician under Sir John Russel at the Rothamsted Experimental Station. He remained at this post until 1933, and it was here that Fisher was faced with some of the practical problems of data analysis. This work at Rothamsted led him, over a span of approximately ten years, to conceptualize and refine the method of analysis of variance. The analysis of variance was nearing its present form by the early 1930's. Many of the further deveopments in the general linear model were made by some of Fisher's proteges at Rothamsted.

We note that Fisher never wrote down linear models in the way we know them today. In contrast to today's text books, it is surprising to learn that The Design of Experiments, which first appeared in 1935, is almost devoid of algebraic equations.

About the time Fisher came on the statistical scene, Karl Pearson was at his height. Statistics was truly in its infancy, and much of the work was in the areas of correlation and curve fitting. In addition, the work was often cast in a more theoretical mold, relying on large sample distribution. Student's (1908a) classic paper, appearing in the climate of the times as it did, was largely ignored by the powerful Pearsonian school. Fisher, however, was later to champion Student's cause and provide the necessary mathematical rigor in proving many of Student's ideas and showing their applicability in more general settings.

3.2. The Development of the Analysis of Variance

After finishing his studies at Cambridge, Fisher spent the years 1915-1919 teaching mathematics at various public high schools. During this time he produced a number of papers, two of which proved to be very important as his career progressed.

Inspired by earlier papers of Soper (1913) and Student (1908a), Fisher (1915)[1] provided the analytic form of the distribution of the correlation coefficient from small samples. His favorite tool of n-dimensional geometry was first used here and was later to be used repeatedly in many of his works. Here he derived the distribution of the correlation coefficient and noted that the distribution was highly skewed (see also Student (1908a)) but that by transforming according to $t = \tan\beta = r/\sqrt{1-r^2}$ and $\tau = \tan\alpha = \rho/\sqrt{1-\rho^2}$ the range of the curve would be extended from $-\infty$ to $+\infty$ and, when $\rho = 0$, the curve approached normality with a form identical to that of Student's Z.

It was in his 1918 paper that Fisher first partitioned genetic variation into its component parts. Box (1978) noted that the original idea for this

---

[1] It is perhaps of historical interest to note that of the nearly 300 papers which Fisher wrote in his lifetime, this was the only one to appear in Biometrika The antagonism between Fisher and Pearson developed shortly after this paper was published, and, with Pearson as editor, this avenue of publication for Fisher became increasingly untenable.

simple algebraic identity, which was to play a major role in Fisher's later works dealing with the Analysis of Variance, was attributable to J. A. Harris.

In calculating dominance ratios (ratios of genetic variation), Fisher (1918) was thinking along the lines of $Z = \log_e \sigma_\alpha^2/\sigma_\tau^2$ rather than $\sigma_\alpha^2/\sigma_\tau^2$. His feeling was that this form had an intrinsic invariance property as a transformation and that the resulting transformed ratio would be approximately normal. The Z transformation was also applied by Fisher (1919) in investigating the resemblence between twins. Fisher (1921a) also applied the Z transformation to inter- and intra-class correlations and was leading toward the ratio $Z = \log_e \frac{1}{2}[(1-\rho)/(1+\rho)]$ recognizable as nearly the form of an F statistic. It was also in this paper that Fisher showed that the distribution of the Z statistic approached normality as the sample size increased.

The first of several applied papers which Fisher authored during his sojourn at Rothamsted appeared in 1921 (Fisher (1921b)). It is through these applied papers and through the early editions of <u>Statistical Methods</u> that one sees the maturing view of the analysis of variance.

One of Fisher's first duties as chief statistician at Rothamsted was to study the massive amounts of data that had accumulated for over 70 years of experimentation of the effects of manural treatments on grain yields. Fisher (1921b) studied the yields from the Broadbalk wheat fields which had been under uniform treatment since 1852 with a view toward ascertaining the principal sources of their variation over time. In plotting the mean yields from several experimental plots, he noticed slow undulating patterns and suggested that the total variation might be attributable to (1) annual variation, (2) deterioration, and (3) other slow changes. Earlier (Fisher (1916)) Fisher showed an interest in Student's ideas on modeling repeated measures data in terms of p^{th} order polynomials (Student (1914)). It was in developing an analysis for these data that Fisher found a use for them. He fitted orthogonal polynomials up to the 5^{th} order and called the linear component the deterioration effect; the 2^{nd} to 5^{th} order terms were attributable to "other slow changes" and annual variation was the error term.[1] Although cast in a regression framework, we recognize this as an early analysis of variance. Fisher's results were arranged rather haphazardly, for example, the p-values for deterioration and slow changes were on separate pages. Perhaps it was in an effort to impose some order on the results that led him to begin tabulating the results as an "Analysis of the Total Variance". Later (Fisher (1934)), he was to make the classic remark regarding the analysis of variance: ". . . the impression that I have formed . . ." is ". . . that the analysis of variance . . . is not a mathematical theorem, but rather a convenient method of arranging the arithmetic.".

The problem of assessing the goodness-of-fit of a regression line was not known in the early 1920's. In presenting a solution, Fisher (1922) did not resort to the Z transformation which would have given the result that is commonly known today. Rather, he treated the Z statistic as a modified chi square statistic. He argued that (1) for p = 2, the best estimate of σ^2 is $\hat{\sigma}^2 = \Sigma(Y_i-\hat{Y}_i)^2/n-2$ which is distributed as a chi square with n-2 degrees of freedom;[2] (2) the statistic $[(a-\alpha)\sqrt{n}]/\sigma$ was approximately normal and independent of the estimate of its error and hence

$$Z = [(a-\alpha)\sqrt{n}]/\hat{\sigma} \tag{2}$$

[1] Kalamkar (1933) performed an identical analysis on the yields of mangolds from Rothamsted.

[2] The concept of "degrees of freedom" first appeared in Fisher's writings in 1922. The author has not been able to trace the exact origin of the term.

was distributed as "Student's Z" which was the form of a Pearson Type IV curve, and which was later shown to be that of an F distribution. The independence of the numerator and denominator in (2) apparently was obvious to Fisher in his 1922 paper; however, he gave an expanded proof of the independence in 1925 (Fisher(1925)). In reference to this latter paper, Seal (1967) suggested that Fisher had not made a detailed study of least squares theory and his proof in this paper of the indepencence of the elements of $\hat{\underset{\sim}{\beta}}$ and $\hat{\sigma}^2$ differed from the proof given by Pizzette (1871) only in his use of n-dimensional geometry instead of Fourier integration theory.

The very general utility of the Z distribution was not fully understood in the early 1920's. In a bench mark paper, Fisher (1928) provided a general treatment of the wide utility of Z. Here he formally defined the Z distribution in terms of the ratio of the standard deviations from two random samples from normal distributions. We note that Fisher's $Z = \frac{1}{2}\log(s_1/s_2)$ is related to the F distribution by a simple transformation. In his discussion he noted how the distribution of Z is dependent upon the sample sizes (degrees of freedom) and showed the relationship of Z to the normal, Student t, and chi square distributions. He also illustrated the use of Z for testing the significance of regression coefficients and multiple correlation coefficients. He laid out a formal analysis of variance for a one-way classification which, except for a modification from Z to F, has remained unchanged. Motivation for the analysis of variance at this stage was in terms of the intraclass correlation model; and it is this form that Fisher used when he introduced the new method of analysis of variance in his first edition of _Statistical Methods_. Fisher's 1928 paper provided the only mathematical basis of the analysis of variance.

Some of the early uses of the analysis of variance method serve to strengthen the view that Fisher treated the analysis as more of an arithmetical tool, rather than from the point of view of linear models as we tend to view this method today.

The earliest application of analysis of variance to factorial-type data was given by Fisher and Mackenzie (1923). The experiment was of the form of a split-plot where 12 plots containing different varieties of potatoes were each split for six different types of fertilizers. Viewed with the benefit of hindsight, the design leaves much to be desired; for example, the varietal plots were not arranged in blocks and the fertilizer treatments, the split-plots, were not randomized within each main plot. The separate sources of error that arise in such a design were not clearly understood at this time, and the sources of variation were partitioned in Fisher and Mackenzie (1923) as:

Source	d.f.
Fertilizers, F	5
Varieties, V	11
F × V	55
Error	141

All effects were tested against the error mean square with 141 d.f. Interestingly, Fisher further considered a multiplicative model and found least squares estimators for the parameters by iterative nonlinear least squares.

Eden and Fisher (1927) gave the analysis for the first example of a 2^3 experiment in randomized blocks. The various sources of variation were partitioned into individual degrees of freedom representing main effects and first and second order interactions "without any complex statistical analysis by simple arithmetical additions and subtractions".

The problems that the analysis of variance encountered in the late 1920's were in the analysis of nonorthogonal experiments. In analyzing a nonorthogonal experiment, Eden and Fisher (1929) only partitioned the orthogonal parts and the nonorthogonal parts were put into the error term. Yates (1933) noted that Fisher thus neglected an interaction of much more consequence than the two interactions orthogonality allowed him to retrieve from the total sum of squares. Seal (1967) provided a more detailed analysis of the Eden and Fisher data.

Beginning in the early 1930's other workers began to express the various sources of variation in the analysis of variance in terms of linear models.

4. The Development of the General Linear Model

4.1. Introduction

By the early 1930's the analysis of variance was beginning to fall into the mold in which it is known today. The analysis of nonorthogonal data still posed a problem at this time, particularly with the introduction of analysis of covariance methods and the analysis of cross classification data having unequal numbers of observations in the subclasses.

The linear models began to formally materialize by 1931. These early models were essentially attempts at writing, in equation form, the various sources of variation in the analysis of variance table.

Atkin (1935) provided the first abstraction of the linear model into its generic matrix algebra form. The concepts of estimable parametric functions were formally introduced by Rao (1945).

Recent authors have advocated a return to a simpler cell means formulation of the general linear model. This formulation is not new and is a reversion to earlier forms of Irwin (1931) and Yates (1934).

4.2. Early Models

Allan and Wishart (1930) introduced the linear model into the statistical literature. They formalized the additive contributions of blocks and treatments in a randomized blocks design by writing down a model for the observations as a linear function of treatment and block "effects". They gave the model as

$$Y = b_p + t_p \quad ,$$

where, in their notation, Y denotes an observation in the p^{th} block and receiving the p^{th} treatment. They later added a constant term to each observation which would correspond to the overall mean and wrote

$$Y = k + b_p + t_p \quad ,$$

By today's standards the above model is glaringly incomplete, since it is missing an error term and is thus divorced from assumptions regarding the underlying sampling structure. This omission was quickly filled by Irwin (1931) who, in considering a two-way cross classification, wrote a model of the form

$$Y_{ijk} = \mu_{ij} + e_{ijk} \quad ,$$

where μ_{ij} represented the population mean corresponding to the ij^{th} cell, and

$$E(e_{ij}) = 0 \ , \qquad Var(e_{ij}) = \sigma^2 \quad .$$

In the same paper Irwin later reparameterized his model by defining row and column effects as deviations from an overall mean, thus giving a model of the form $Y_{ijk} = b_i + t_j + e_{ijk}$; however, he carefully contented himself with estimating only $b_i + t_j$ rather than b_i or t_j.

The extension of the analysis of variance to cover the case of unequal frequencies of a two-way classification was considered by Brandt (1933) who apparently did not realize all the complications that the absence of orthogonality entails, since he incorrectly assumed that the various sums of squares in the analysis of variance table must be additive as they would be in an orthogonal experiment.

Brandt's results were corrected by Yates (1934) who presented a detailed acoount of the methods of analysis of a two-way table in a nonorthogonal setting. Yates (1934) initially considered a model of the form

$$Y_{ijk} = \mu_{ij} + e_{ijk} \quad , \tag{3}$$

where

$$e_{ijk} \quad \text{i.i.d.} \quad N(0,\sigma^2) \quad .$$

He then defined

$$\mu = \Sigma\Sigma\mu_{ij}/ab \quad ,$$

$$\alpha_i = \bar{\mu}_{i.} - \mu \quad , \tag{4}$$

and

$$\beta_j = \bar{\mu}_{.j} - \mu \quad .$$

The model could then be reparameterized as

$$Y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk} \quad , \tag{5}$$

which, by way of (4) also included the restrictions

$$\alpha_. = 0 \ , \quad \beta_. = 0 \quad . \tag{6}$$

This paper of Yates' is truly a classic and has stood up well over time. Today several papers are being published which attempt to develop a better understanding of the analysis of nonorthogonal data with particular reference to the structure of the various hypotheses which can be tested for models like (5). In doing so there is frequent reference to Yates (1934).

4.3. Two Important Generalizations

Aitken (1935) extended the least squares theory to include the case of correlated errors. In doing so he wrote down for the first time the general linear model in compact matrix form. Allowing a slight change in notation into more familiar terms, he wrote a general model as

$$\mathbf{y} = \mathbf{X}\mathbf{b} \tag{7}$$

under the general case of

$$\mathrm{Var}(\mathbf{y}) = \mathbf{V} \quad . \tag{8}$$

He showed that by minimizing $(\mathbf{y}-\mathbf{X}\mathbf{b})\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\mathbf{b})$ with respect to the elements of $\mathbf{b}$ one obtained normal equations of the form

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

which had a solution

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad . \tag{9}$$

Estimators of the form of (9) are commonly referred to today as Aitken estimators in honor of their developer.

Note that Aitken did not explicitly consider a vector of error terms in (7), but he did so implicitly by way of (8) and his consideration of minimization of the weighted inner product of the residual vector $\mathbf{y} - \mathbf{X}\mathbf{b}$. Aitken, like others of his time, was not careful to distingish between the assumed model and the assumed underlying error structure. While Aitken would specify a model as in (7) and (8), today we would write the model as

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{b} + \underset{\sim}{e}$$

for

$$\underset{\sim}{e} = \underset{\sim}{y} - E(\underset{\sim}{y}) = \underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{b}$$

and where

$$\underset{\sim}{e} \sim (0,V) \quad .$$

Rao (1945) formally introduced the concepts of estimable functions. He distinguished between two types of functions — estimating functions and error functions. A linear function $\underset{\sim}{t}'\underset{\sim}{y}$ was said to belong to the error space if $E(\underset{\sim}{t}'\underset{\sim}{y}) = 0$, otherwise, it belonged to the estimation space. The totality of all linearly independent error functions consistuted the error space. The space orthogonal to the error space was the estimation space, which in turn was the orthogonal projection of $\underset{\sim}{y}$. The best linear unbiased parametric functions (BLUEs) of estimable functions consisted of scalar products of the vector $\underset{\sim}{y}$ with vectors lying in the estimation space. In the generic setup, $\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{b} + \underset{\sim}{e}$, a parametric function $\underset{\sim}{t}'\underset{\sim}{b}$ is estimable if and only if $\underset{\sim}{t}' = \underset{\sim}{q}'\underset{\sim}{X}$ for some $\underset{\sim}{q}'$. Rao (1945) provided the generalization of the Gauss-Markov theorem for a model of the form $\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{b} + \underset{\sim}{e}$: if $\underset{\sim}{t}'\underset{\sim}{b}$ is an estimable function then the BLUE of $\underset{\sim}{t}'\underset{\sim}{b}$ is $\underset{\sim}{t}'\underset{\sim}{b}^o$ where $\underset{\sim}{b}^o$ is any solution to the normal equations.

4.4. Discussion

Rao's contribution allowed an important generalization of linear models. Fixed effects representing contributions of rows, columns, treatments, blocks, etc. in experimental design work were allowed to exist independently of deviations from means. One could now consider models of the form of (5) without requiring restrictions such as (6). Thus there are two models[1] for a two-way classification such as Yates (1934) considered. These models would be of the form:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk} \tag{10}$$

and

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk} \tag{11}$$

where

$$\Sigma\alpha_i = 0 \quad \text{and} \quad \Sigma\beta_j = 0 \quad .$$

Rao's ideas on estimable functions allowed consideration of models like (10) where the nature or structure of the fixed effects could remain unspecified. Models like (11) were used by Irwin (1931) and Yates (1934) who carefully noted the origin of (11) by way of (3) and (4). Experimenters such as Yates and Irwin would have little use for the concept of estimability.

The nature of the difference between (10) and (11) may appear superficial, but the concept of estimability makes a strong distinction. While all effects and any linear combination of the effects in (11) are estimable, such is not the case for (10). Individual effects in (10) are not estimable — only certain contrasts; i.e., those contrasts which can be expressed as a linear combination of the expected values of the observations. Parametric functions which are estimable in (10) will, of course, be estimable in (11); but the reverse is not always true.

Concepts that surround the various hypotheses that can be tested about the parameters in models such as (10) are much too detailed to go into here. Excellent coverage is given in Searle (1971). Recently, there has been a surge of

---

[1] Many other models involving restrictions on the parameters are, of course, also possible. We shall consider only two.

interest in understanding the nature of hypotheses tested by the usual least squares or fitting constants methods. Many of these hypotheses involve complicated linear combinations of the α_i's and β_j's in (10). Further complications arise when one considers interaction models and testing hypotheses for data with missing subclasses. To a large extent, the recent work of Hocking, et al. (1975) takes aim at this confusion and considers linear models cast in the form of (3). Much of their work has been involved with explaining least squares hypothesis tests in terms of their explicit forms as weighted linear combinations of the population cell means. Models of the form of (3) are now known as "μ_{ij}-models" or "cell means models". It is a reversion to the ideas of Irwin (1931) and Yates (1934), and it is perhaps what Fisher had in mind when he formulated the analysis of variance, although his lack of model specification makes such speculation difficult.

Thus the linear model has made an evolutionary full swing in returning to its simpler origins. The μ_{ij} models today differ from those in the early 1930's in the depth and breadth of the theoretical structure which has evolved along with the general linear model. Ideas regarding error structure, inference, and power, concepts of comparison-wise and experiment-wise error rates have evolved along with the linear model and serve to strengthen and broaden the foundation of a model as deceptively simple as $E(Y_{ijk}) = \mu_{ij}$.

5. References

Aitken, A. C. (1935). On least squares and linear combination of observations. Proc. R. Soc. Edinb. A 55:42-48.

Allan, F. E. and Wishart, J. (1930). A method of estimating the yield of a missing plot in experimental work. J. Agri. Sci. 20:399-406.

Box, J. F. (1978). R. A. Fisher: The Life of a Scientist. John Wiley and Sons., Inc., New York.

Brandt, A. E. (1933). The analysis of variance in a 2 x s table with disproportionate frequencies. JASA 28:164-

Bravais, A. (1846). Sur les probabilités des frreurs de situation d'un point. Mem. Acad. Roy. Sci. Inst. France 9:255-332.

Eden, T. and Fisher, R. A. (1927). Studies in crop variation. IV. The experimental determination of the value of top dressings with cereals. J. Agric. Sci. 17:548-562.

Eden, T. and Fisher, R. A. (1929). Studies in crop variation. VI. Experiments on the response of the potato to potash and nitrogen. J. Agric. Sci. 19: 201-213.

Edgeworth (1892). Correlated averages. Phil Mag. 190-204.

Eisenhart, C. (1964). The meaning of "least" in least squares. J. Wash. Acad. Sci. 54:24-33.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika 10:507-521.

Fisher, R. A. (1916). "Biometrika". Eugenics Review 8:62-64.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinb. 52:399-433.

Fisher, R. A. (1921). Studies in crop variation. I. An examination of the yield of dressed grain from Broadbalk. J. Agric. Sci. 11:107-135.

Fisher, R. A. (1922). The goodness of fit of regression formulae, and the distribution of regression coefficients. J.R.S.S. 85:597-612.

Fisher, R. A. (1925). Applications of "Student's" distribution. Metron 5(3): 90-104.

Fisher, R. A. (1928). On a distribution yielding the error functions of several well known statistics. Proc. Int. Congr. Math., Toronto (1924), 805-813.

Fisher, R. A. (1934). Discussion of "Statistics in Agricultural Research" by J. Wishart. J.R.S.S. Suppl. 1:51-53.

Fisher, R. A. (1935). The Design of Experiments. Oliver and Boyd, London.

Fisher, R. A. and Mackenzie, W. A. (1923). Studies in crop variation. II. The manurial response of different potato varieties. J. Agric. Sci. 13:311-320.

Galton, F. A. (1877). Typical laws of heredity in man. Lecture delivered before the Royal Institution of Great Britian, February 9, 1877.

Galton, F. A. (1885). Lecture to the Anthropological Institution.

Galton, F. A. (1888). Correlation and their measurement chiefly from anthropometric data. Proc. Royal Society 45:135-145.

Galton, F. A. (1889). Natural Inheritance. Macmillan and Company, London.

Gosset, W. S. (1908a). The probable error of a mean. Biometrika 6:1-25.

Gosset, W. S. (1908b). Probable error of a correlation coefficient. Biometrika 6:302-310.

Gosset, W. S. (1914). The elimination of spurious correlation due to position in time or space. Biometrika 10:179-180.

Hocking, R. R. and Speed, F. M. (1975). A full rank analysis of some linear model problems. JASA 70:706-712.

Irwin, J. O. (1931). Mathematical theorems involved in the analysis of variance. J. Royal Stat. Soc. 94:284-300.

Kalamkar, R. J. (1933). A statistical examination of the yields of Mangolds from Barnfield at Rothamsted. J. Agri. Sci. 23:161-175.

Pearson, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. Phil. Trans. A 187:253-318.

Pearson, K. P. (1920). Notes on the history of correlation. Biometrika 13:25-45.

Pearson, K. P. (1930). The Life, Letters and Labours of Francis Galton. Vol. III.a. Correlation, Personal Identification and Eugenics. Cambridge University Press, Cambridge.

Pizzetti, P. (1891). I fondamendi matematici per la critica dei risultati sperimentali. Genova.

Plackett, R. L. (1972). The discovery of the method of least squares. Biometrika 59:239-251.

Rao, C. R. (1945). Generalization of Markoff's theorem and tests of linear hypotheses. Sankhyā 7:9-16.

Seal, H. L. (1967). The historical development of the Gauss linear model. Biometrika 54:1-24.

Searle, S. R. (1971). Linear Models. John Wiley and Sons, Inc., New York.

Soper, H. E. (1913).
Biometrika 10:507-521.

Speed, F. M. and Hocking, R. R. (1976). The use of the R( ) notation with unbalanced data. The Amer. Stat. 30:30-34.

Speed, F. M., Hocking, R. R. and Hackney, O. P. (1978). Methods of analysis of linear models with unbalanced data. JASA 73:105-112.

Urquhart, N. S. and Weeks, D. L. (1979). Linear models in messy data: some problems and alternatives. Biometrics 34:696-705.

Urquhart, N. S., Weeks, D. L. and Henderson, C. R. (1969). Estimation associated with linear models: A revisitation. Paper Number BU-195, Biometrics Unit, Cornell University, Ithaca, New York.

Weldon, (1890). The variations occurring in certain Decapod Crustacea — I Crangonvulga. Proc. Royal Soc. 47:445.

Weldon, (1893). On certain correlated variations in Carcinus moenas. Proc. Royal Soc. 54:318-329.

Yates, F. (1933). The principle of orthogonality and confounding in replicated experiments. J. Agri. Sci. 23:108-145.

Yates, F. (1934). The analysis of multiple classifications with unequal numbers in the different classes. J. Amer. Stat. Assoc. 29:51-66.

Yule, G. U. (1897). On the theory of correlation. J. Stat. Soc. 60:812-854.